

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-220689

(43)公開日 平成11年(1999)8月10日

(51)Int.Cl.⁶

H 04 N 5/91
G 06 T 13/00

識別記号

F I

H 04 N 5/91
G 06 F 15/62

N
3 4 0 A

審査請求 未請求 請求項の数4 FD (全15頁)

(21)出願番号 特願平10-33984

(22)出願日 平成10年(1998)1月31日

(71)出願人 591015854

株式会社メディア・リンク・システム
東京都中央区東日本橋2-2-10 東日本
橋オリモビル

(72)発明者 福田 橘

東京都中央区東日本橋2-2-10 株式会
社メディア・リンク・システム内

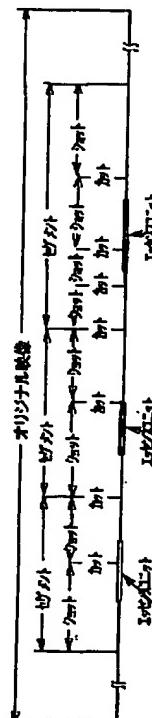
(74)代理人 弁理士 松井 晃一

(54)【発明の名称】 映像ソフト処理装置及び同処理プログラム記録記憶媒体

(57)【要約】

【課題】 別の視点からの要約生成を可能にし、また視聴者の都合に合わせ、要約再生時間を自由に設定可能な映像ソフト処理装置及び同処理プログラム記録記憶媒体を提供。

【解決手段】 映像の各カット位置検出。各カット位置から次のカット位置までの各ショットについて、持続時間所定値以下のもの、セリフ、文字ほか注目映像ソフト構成要素が両者間に跨がっているものを連結。単数又は複数のショットから成る各セグメント形成。効果音、文字ほかの注目映像ソフト構成要素の存在状態に対応し、各セグメントの中から所要長さの映像をエッセンスユニットとして取り出し(請求項1, 3)。更に、その抽出動機の映像ソフト構成要素存在状態に対応して各エッセンスユニットに評価点付与。これを基に指定再生時間になるようにエッセンスユニットの幾つか選択、これらを再生(請求項2, 4)。



【特許請求の範囲】

【請求項1】 映像の各カット位置を検出するカット位置検出手段と、該検出された各カット位置から次のカット位置までの各ショットについて、その持続時間が所定値以下のもの、或いは効果音、文字、セリフその他注目した映像ソフト構成要素が両者間に跨がっているものを連結し、単数又は複数のショットから成る各セグメントを形成するセグメント形成手段と、効果音、文字、セリフ、対象物その他注目した映像ソフト構成要素の存在状態に対応し、各セグメントの中から所要長さの映像をエッセンスユニットとして抽出するエッセンスユニット抽出手段とを備えたことを特徴とする映像ソフト処理装置。

【請求項2】 前記各エッセンスユニットについて、その抽出動機とされた映像ソフト構成要素の存在状態に対応した評価点を付与する評価手段と、前記評価点に基づき指定された再生時間になるように前記エッセンスユニットの中の幾つかを選択するエッセンスユニット選択手段とを更に備えることを特徴とする請求項1に記載の映像ソフト処理装置。

【請求項3】 映像の各カット位置を検出するカット位置検出手処理と、該検出された各カット位置から次のカット位置までの各ショットについて、その持続時間が所定値以下のもの、或いは効果音、文字、セリフその他注目した映像ソフト構成要素が両者間に跨がっているものを連結し、単数又は複数のショットから成る各セグメントを形成するセグメント形成処理と、効果音、文字、セリフ、対象物その他注目した映像ソフト構成要素の存在状態に対応し、各セグメントの中から所要長さの映像をエッセンスユニットとして抽出するエッセンスユニット抽出処理とをコンピュータ上で実行する映像ソフト処理プログラムを記録したことを特徴とする記憶媒体。

【請求項4】 前記各エッセンスユニットについて、その抽出動機とされた映像ソフト構成要素の存在状態に対応した評価点を付与する評価処理と、前記評価点に基づき指定された再生時間になるように前記エッセンスユニットの中の幾つかを選択するエッセンスユニット選択処理とが更に含まれることを特徴とする請求項3に記載の映像ソフト処理プログラムを記録した記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は映像ソフト処理装置及び同処理プログラム記録記憶媒体に関し、詳しくは、各カット位置から次のカット位置までの各ショットについて、その持続時間が所定値以下のもの、或いは、効果音、文字（テロップ）、セリフその他注目した映像ソフト構成要素が両者間で跨がっているものなどを連結してセグメントとし、これら注目した構成要素の存在状態に対応し、これらセグメントの中から所要長さの映像をエッセンスユニットとして抽出し、これを視聴すること

で、夫々の映像ソフトの内容を、視聴者が短時間で容易に感得出来るようにした映像ソフト処理装置及び同処理プログラム記録記憶媒体に関する。なお、「映像ソフト」とは「映像ソフトウェア」を略して表記しているものである。

【0002】

【従来の技術】 ビデオ、DVD、衛星放送等、映像ソフトがあふれている世界では、見たい映像を早く探し、それを短時間で見るというニーズが出てくる。（ビデオ＝ビデオテープ或いはビデオテープレコーダー。DVD＝デジタルバーサタイルディスク。）

しかし、従来のビデオテープレコーダ、DVDプレーヤ等では、録画した映像を時間節約して見る方法と言えば、早送りぐらいためしか方法がなかった。そこで出願人は、平成9年特許願第304862号などにより、映像ソフトから要約を抽出する装置等を提案した。要約は映像ソフトの要点を端的に表わすもので、視聴者はこの要約を見ることで、通常なら長時間掛かる映像ソフトの内容把握を、極めて短時間のうちに済ますことができる。

【0003】 これら発明では、音声その他映像ソフトの構成要素の状態変化に着目して、夫々のショットから1枚または複数枚のフレーム（画像）を代表画像として取り出し、これらを見ることで、その映像ソフトの内容を視聴者が短時間で感得出来るようにしている。（なお、上記出願では、映像の切り替わり位置から次の切り替わり位置までの一連のフレームを「カット」と称したが、これは「ショット」と称する方が的確と思料される。本明細書ではこちらの呼称を使用する。）

【0004】

【発明が解決しようとする課題】 上記発明は、同じ短時間でも、単なる早送りに比べ映像ソフトの内容を一層的確に把握出来るという新たな効果をもたらした。しかし、技術は常に改善が求められる。このため、先の発明に関し更に研究を進めていたところ、下記の如き課題が見つかった。

【0005】 第1は、要約の更なる質の向上である。即ち、先の発明ではショットを基礎にして、そこから代表画像を抽出していた。しかし、映像ソフトのジャンルは多岐に亘る。この為、生成された要約にいさか冗長度が感じられるなど、ショットという概念だけが要約生成の鍵の全てではないと感じられる面もあり、より多様性を持たせるため、別の観点からの映像ソフト要約手法の出現が望まれた。

【0006】 第2は、再生時間を自由に設定可能にすることである。視聴者は一人一人生活様式が違う。要約を見るのに当たられる時間も一人一人異なる。従って短時間で、という本発明の目的を突き詰めるなら、これら要約を見るに要する時間（再生時間）に関しても、視聴者夫々の事情に応じ、自由に設定可能であることが望ましい。

【0007】本発明の目的は、これら新たに出現した課題を解決することにあり、別の視点からの要約生成を可能にし、また、視聴者の都合に合わせ、要約再生時間を自由に設定可能である映像ソフト処理装置及び同処理プログラム記録記憶媒体を提供することにある。

【0008】

【課題を解決するための手段】上記目的達成のため請求項1の発明では、映像の各カット位置を検出するカット位置検出手段と、該検出された各カット位置から次のカット位置までの各ショットについて、その持続時間が所定値以下のもの、或いは効果音、文字、セリフその他注目した映像ソフト構成要素が両者間に跨がっているものを連結し、単数又は複数のショットから成る各セグメントを形成するセグメント形成手段と、効果音、文字、セリフ、対象物その他注目した映像ソフト構成要素の存在状態に対応し、各セグメントの中から所要長さの映像をエッセンスユニットとして抽出するエッセンスユニット抽出手段とを備える。

【0009】また、請求項2の発明では、請求項1の構成に加え、各エッセンスユニットについて、その抽出動機とされた映像ソフト構成要素の存在状態に対応した評価点を付与する評価手段と、前記評価点に基づき指定された再生時間になるように前記エッセンスユニットの中の幾つかを選択するエッセンスユニット選択手段とを更に備える。

【0010】また、請求項3の発明では、映像の各カット位置を検出するカット位置検出手段と、該検出された各カット位置から次のカット位置までの各ショットについて、その持続時間が所定値以下のもの、或いは効果音、文字、セリフその他注目した映像ソフト構成要素が両者間に跨がっているものを連結し、単数又は複数のショットから成る各セグメントを形成するセグメント形成手段と、効果音、文字、セリフ、対象物その他注目した映像ソフト構成要素の存在状態に対応し、各セグメントの中から所要長さの映像をエッセンスユニットとして抽出するエッセンスユニット抽出手段とを実行する。

【0011】また、請求項4の発明では、請求項3の処理に加え、前記各エッセンスユニットについて、その抽出動機とされた映像ソフト構成要素の存在状態に対応した評価点を付与する評価手段と、前記評価点に基づき指定された再生時間になるように前記エッセンスユニットの中の幾つかを選択するエッセンスユニット選択手段とを更に実行する。

【0012】

【発明の実施の形態】以下、本発明の詳細を図示実施の形態例に基いて説明する。始めに用語について説明する。

1. 用語の説明。

・オリジナル映像（図1参照。以下同じ。）

テレビ放送、ビデオ映像をデジタル録画した全体の映

像。

・カット

カメラの切り替え点、連続した映像の切れ目。

・ショット

カットからカットまでの映像。

・セグメント

ショットをつないで区切りのいい長さにした映像。始めのセグメントから終わりのセグメントまで連続したものがオリジナル映像。

・セグメントリスト（表1）

セグメントを管理するための表。VE（要約）とオリジナル映像の切替操作、オリジナル映像をスキップ送りしてみる場合などに用いる。

【0013】・エッセンスユニット

セグメントの中で特に重要だと判定される映像。効果音、文字、セリフ（人の声）等のAV要素（映像ソフトの構成要素）が存在するか否かを基に判定する。

・エッセンスユニット評価

各エッセンスユニットをAV要素の種類、長さ（持続時間）、全体構成等により採点する。

・エッセンスユニットリスト（表2）

エッセンスユニットを管理するための表。

・セグメント／エッセンスユニット対応リスト

セグメントNO.と、対応するエッセンスユニットNO.を記録したもの。

【0014】・インデックス画面（図12）

各セグメントの始めの画像をサムネイル化（小型化）し、そのサムネイル画像を複数個、1画面に張り付けた画面。サムネイル画像と該当するオリジナル映像にリンクを張っており、サムネイル画像で指定したオリジナルを再生できる。

・映像の構成（図2）

映像はタイトル、オープニング、中間部、エンディングにより構成される。

・タイトル

映像の題名。VEにおいてタイトルは重要であり、作成したVEを再生したときに最初にタイトルが表示されるようにする（図10）。

・オープニング（図2）

映像スタート後の導入部。コンテンツの概要や全体の把握をする上で重要である。

・中間部（図2）

中間部では、視聴者を映像の世界に引き込み、その内容を把握してもらえるようにストーリーが構成され、カメラワーク、セリフ、BGM、テロップ等の要素が盛り込まれる。中間部は絞り込んでいく部分や盛り上がりのある部分いろいろなシーンによって構成される。

・エンディング（図2）

オリジナルビデオは一般的に最後の数シーンにまとめる。この中で結末への導入と結論、最後のEND/PIN

E／終、の文字は重要である。

・AV要素

AV要素の中では、効果音、テロップ、対象物、セリフの有無が重要と解される（図3データ蓄積部19内「要素データ」）。

【0015】2. 次に実施の形態例のVEプレーヤ100の構成を説明する（VE=ビデオエッセンス）。図3に於て、11はTV受信部であり（TV=テレビジョン）、視聴者が選択した1又は複数局のTV放送を受信する。受信データは入力制御部12でデジタル化され、映像ソフト蓄積部13（メモリ、ハードディスク等）に蓄積される。14はビデオテープデッキであり、ここで再生された映像信号及び音声信号も入力制御部12でデジタル化され、映像ソフト蓄積部13に蓄積される。

【0016】16は光ディスクプレーヤであり、視聴者が所望するDVD、レーザその他の光ディスクのデータを読み出す。TV放送やビデオテープの信号をデジタル化したのは、後の処理及び再生の際の読み出しを容易にするためである。従ってDVDのように既にデジタル化されているものについてはデジタル化する必要はなく、後の処理の際、直接その媒体からデータが読み出される。17はインターネット接続部であり、ユーザが所望するホームページ等から映像ソフトを受信し、映像ソフト蓄積部13に蓄積する。なおインターネットから取得するものもデジタル化されているので、そのまま映像ソフト蓄積部13に蓄積される。

【0017】18はカット判別部である。このカット判別部は、光ディスク（16）或いは映像ソフト蓄積部13から読み出される映像ソフトの映像の変化状態に着目し、夫々のカット位置を検出する。このカット判別部18が請求項にいうカット位置検出手段又はカット位置検出処理を実行する手段に当たる。カット位置は、例えば、上記平成9年特許願第304862号に記載の手法や、書籍「ビデオデータベースシステム」などに開示された手法で判別できる。なお、書籍「ビデオデータベースシステム」の原書名は「VIDEO DATABASE SYSTEMS(Issues, Product, and Applications)」、著者は、Ahmed K.E. Imagarmidほかであり、出版は、Kluwer Academic Publishers、出版社所在地は、米国マサチューセッツ州ノーウェル市である。

【0018】詳細はこれらの記載を引用するが、理解を容易にするため平成9年特許願第304862号に記載の手法の一例をかいつまんで以下に記載する。

1) ピクセルの集約処理

画像は2次元的な広がりとして例えば 640×400 というような点（ピクセル）の集まりで表現される。テレビ映像ではこのような画像が1秒間に30枚供給される。処理を簡単にするため、先ずはこのピクセルの集約化を行う。例えば 4×4 の点で集約すれば、原映像は $1 / 16$ のデータ量にすることができる、 8×8 で取り出せ

ば、 $1 / 64$ に集約できる。こうすると、ズームアップ／ズームダウンのときや、ゆっくりと画像が変化するとき、この変化を捨象する働きがある。

【0019】このような処理を行った各点について、 $a(t, x, y)$

をとりだす。ここに、 t ：時刻 x, y ：集約した映像の座標 a ：その点 (x, y) の色の値であり、 $a = R + G + B$ などとするのが一般的である（R, G, Bは3原色情報の値）。

【0020】2) 時間方向の集約化処理

時間的に変化する映像データは、そのまま扱うには冗長であり、またムダも多い。そこで、前記処理を行なった各点のデータ $a(t, x, y)$ に対して、例えば映像の時間的な差分比較処理をする。これは、上記処理後の各点について、時間的な変化を見るもので、

$$d(t, x, y) = a(t, x, y) - a(t - \Delta t, x, y)$$

で求められる。ここに、 Δt は適宜の時間幅である。

【0021】これを画像全体 (x, y) について集計する。式で表わせば、

【数1】

$$D_a(t) = \sum_{x, y} (d(t, x, y))$$

である。これは、時間的に Δt だけ隣り合う2枚の画像の間の変化量（差分）を示している。ここで値が大きい順に、 $D_a(t)$ をN個とりだす。このようにして取り出した $D_a(t)$ の例を図13に示す。値が大きいところ、例えばCS1, CS2は、そこで画像に大きな変化がある、即ちそこがカット位置である可能性が高い。従って、このCS1とCS2をカットとして、その間を一つのショットとする。このようにして検出された各ショットのデータは、データ蓄積部19に蓄積される（図3データ蓄積部19内「ショットデータ」）。

【0022】再び図3に於て、21は要素検出部であり、処理対象映像ソフトの各要素、例えば効果音、テロップ、セリフ、対象物、音楽その他の要素の存在状態を検出する。これらのデータはデータ蓄積部19に蓄積される（データ蓄積部19内「要素データ」）。これら要素の検出は、映像ソフトを蓄積する際に行ない、映像ソフトと一緒に蓄積しておくと良い。これらは、後述するセグメントの形成、エッセンスユニットの抽出、エッセンスユニットの評価などに利用されるので、こうしておくことで、後の処理が早くなる。

【0023】なお、セリフは例えばスペクトラムを参照して存在を確認する。また人の話声は言葉と言葉の間に通常0.5秒位のブランクがあるので、これも判断要素に使える。対象物の検出は、例えば次のように行なう。先ず、各画像を例えば縦横3つづつの9個の領域に分ける。次いで、夫々の領域に属する各ピクセルについて、その画像成分の分布、例えば各ピクセルの輝度分布を検

出する。輝度は例えば10段階に分ける。この結果、例えば各領域のピクセル数は28,444個であり、そのうち、画面中央の領域では、一番輝度の低い段階のピクセルが1,000個、その次のものが1,500個、…というように、夫々の領域のピクセルの画像成分の分布が求められる。

【0024】次いで、中央の領域の画像成分の分布と、周辺の領域の画像成分の分布の差分を求める。求められた差分の例を図14に示す。縦軸はこの差分の百分率、横軸は、映像ソフトの先頭からのフレーム番号であり、差分は、各フレームについて求めている。この事例の場合、概ね差分が60%（一点鎖線で表示）を越えている部分では、画像の中に人物その他の対象物、即ち、その映像ソフトの内容を端的に表わすものが存在した。

【0025】従って、差分の大きさを適宜に設定することで、画像の中の何らかの対象物の存在を検出することが出来る。この手法で対象物として検出された具体的なものとしては、①背景は部屋の中で、画面中央にパソコンディスプレイが映っている、②背景はビルで、画面中央に人物が映っている、③背景は景色で、画面中央にトンネルを出て来る電車が映っている、などであった。

【0026】22はセグメント形成部で、前記各ショットデータに基き、その持続時間はどのくらいか、二つのショットの間で連続する効果音、テロップ、セリフが存在するかなどを確認し、单一または複数個のショットから成るセグメントを形成する。そのデータはデータ蓄積部19に蓄積される（データ蓄積部19内「セグメントリスト」。表1）。このセグメント形成部22が請求項にいうセグメント形成手段またはセグメント形成処理を実行する手段に当たる。23はエッセンスユニット抽出部で、上記判別（形成）されたセグメントのデータに基き、エッセンスユニットの抽出を実行する。このデータもデータ蓄積部19に蓄積される（データ蓄積部19内「エッセンスユニットリスト」、「セグメント／エッセンスユニット対応リスト」）。このエッセンスユニット抽出部23が請求項にいうエッセンスユニット抽出手段又はエッセンスユニット抽出処理を実行する手段に当たる。

【0027】24はエッセンスユニット評価部で、上記抽出された各エッセンスユニットに付随する効果音その他の映像ソフト構成要素の存在状態を評価し、これらエッセンスユニットについて、言わば点数付けを実行する。このデータもデータ蓄積部19に蓄積される（データ蓄積部19内「エッセンスユニット評価リスト（表2）」）。このエッセンスユニット評価部24が請求項にいう評価手段又は評価処理を実行する手段に当たる。26はユーザインタフェースで、不図示キーボード、マウス及びディスプレイ27上に配置される各ボタン（図11「HOME」ほか）を備えており、視聴者はこれらを操作して所望する要約映像（エッセンスユニッ

ト及び関連する音声データの集合）を視聴する。

【0028】再生制御部28は、これら視聴者の操作に応動し、所望された再生時間に合致するエッセンスユニットの選択とその再生（ディスプレイ27やスピーカ29への出力）、オリジナル映像ソフトの再生、エッセンスユニット／オリジナル映像の切替表示、ライブラリーパネル（図10）の表示、インデックス画面（図12）の表示などを実行する。この再生制御部28が、請求項にいう選択手段又は選択処理を実行する手段に当たる。

【0029】3. VE生成プロセス

次に、VE生成プロセスについて説明する。なお括弧書きは、その処理を主として実行する構成を表わす。

（1）デジタル録画（入力制御部12）

- ・上記図3を引用した説明で述べたとおり、先ずテレビ放送、ビデオ映像（ビデオテープ等からの映像）をデジタル化し、フレーム番号をとり、フレーム番号を指定することによりランダムアクセス可能なファイルとして映像ソフト蓄積部13に保存する。

【0030】（2）カット判別、セグメント（映像の区切り）形成。

- ・カットを判別（検出）し、カットからカットまでをショットとする（カット判別部18）。

- ・1秒以下のショットSSはその前のセグメントに付ける（図5。セグメント形成部）。なお1秒には限定されない。またこの「1秒」が請求項にいう「所定値」に当たる。）

- ・セリフ、テロップがカットをまたがっている場合、そのカット前後の2つのショットを繋ぐ（図5。セグメント形成部22）。

【0031】（3）セグメントリスト

- ・セグメントが決定したら、通しNO.を付け、各セグメントの開始／終了フレームNO.をとり、セグメントリストを作成する（セグメント形成部22）。

- ・セグメントリストはセグメントNO.、セグメント開始フレームNO.、セグメント終了フレームNO.などで構成される（表1）。

【0032】（4）エッセンスユニットの抽出（エッセンスユニット抽出部）

（定義）

- ・エッセンスユニットはセグメントの主要部を抜き出して作られる。

- ・1セグメントより1エッセンスユニットを作る。

- ・エッセンスユニット切り出しの優先順は、効果音、文字（テロップ）、セリフ（人の声）とする。

【0033】（プロセス）

- ・そのセグメントに効果音があった場合で、その区間（T1）+前後1.5秒に文字及びセリフがない場合、その区間（T1）+3秒間、合計T2の区間をエッセンスユニットとして切り出す（図4）。ここでは、所定しきい値THLを越えた音声信号（T1の部分）を効果音

としている。

- ・前後の3秒間の間に、文字、セリフが出てきた場合、文字、セリフが切れるまでエッセンスユニットをのばす。

【0034】・セグメントに効果音がない場合、文字が存在する部分をエッセンスユニットとして切り出す。

・文字の部分にセリフが出てきた場合、セリフが切れるまでエッセンスユニットをのばす。

・セグメントに効果音、文字が無い場合、セリフの部分をエッセンスユニットとして切り出す。

【0035】・同一セグメントに複数の効果音、文字、又はセリフがある場合、最初の効果音、文字、セリフの部分を優先させる。

・セグメントに効果音、文字、セリフがない場合、例えば最初から2秒間をエッセンスユニットとする。これは、セグメント化するに当たり、短いショットを前のショットに連結することとしたので、セグメントの後の方から抽出すると、この短いショット部分から抽出してしまう可能性があり、これでは、ショットが短くて見にくい心配があるからである。

【0036】・エッセンスユニットを抜き出したら、最初の0.5秒は徐々に音量を大きくしていき、終わりの0.5秒間は徐々に絞っていく。こうしておくと、後でエッセンスユニットとエッセンスユニットとを繋いだとき、そこで音楽がブツツリと切れる感じが無くなって、スムーズに音が繋がる。

【0037】(5) エッセンスユニットの採点(評価)
・採点要素は例えば下記のものとする。

効果音 = 10点

文字 = 2点

セリフ = 2点

対象物 = 1点

音楽 = 1点

・各要素の長さはフレーム数で表わす。

・採点方法は各採点要素×フレーム数 の合計とする。

・全体構成による採点アップ

始まりから10%部迄をオープニング部として、その部分の合計点数を10%アップする(図2)。最後の5%部をエンディング部として、その部分の合計点数を10%アップする(図2)。

【0038】(6) エッセンスユニットリストの内容
・エッセンスユニットNo.

スタートからエンド迄エッセンスユニットの通し番号

・エッセンスユニットポイント

各エッセンスユニットの採点

・エッセンスユニットタイム

各エッセンスユニットのフレーム数

【0039】(7) セグメント/エッセンスユニット対応表

・セグメントNO. とエッセンスユニットNO. から成

る。

(8) エッセンスユニットの選択

- ・VE再生時に視聴者がプレイ(再生)時間(又は要約率)を設定する。

・再生制御部28は、それらの再生時間の合計が設定されたプレイ時間に合うように、エッセンスユニットポイントの高い順にエッセンスユニットを選択する。

(9) エッセンスユニット接続・VEプレイ(要約再生)

・選択されたエッセンスユニットを順に接続し、ディスプレイ27上に再生する。

・再生中に、オリジナル-VE間のリンクができるよう、NO.管理を行う。

【0040】(10) インデックス画面

・各セグメントの始めの画像をサムネイル化(小型化)し、そのサムネイル画像を1画面に複数個(例:横5枚×縦4枚)張り付ける(図12)。

・サムネイル画像と該当するオリジナル映像(セグメント)にリンクを張る。

・サムネイル画像がクリックされたときそれとリンクしているオリジナル映像を再生する。

【0041】4. 図示したVEプレーヤの構成を纏めると以下のとおりになる。

(1) 基本機能

・映像フレーム毎にランダムアクセス可能なデジタル映像を記録する。

・VEの再生時間を設定(時間又は圧縮率を指定)しVEを再生する。

・VEを見ているときにボタン一つでオリジナル映像に切り替えられる。

・オリジナル映像からVEへの切り替え。

・指定位置からのVEプレイ。

・単位映像(セグメント又はエッセンスユニット)毎にスキップ進み/戻り(連続操作によりばらばらめくり)操作。

・各エッセンスユニットの第1フレームをサムネイル化し、サムネイル画像を複数表示したインデックス画面(映像目次)表示、及びサムネイル画面を指定しそれとリンクしたオリジナル映像の再生。

【0042】(2) 録画・VE生成処理

テレビ放送、ビデオテープの映像はアナログで録画されているので、その映像をデジタル録画する。デジタル録画された映像はフレームNO.でアクセスできるようなファイルとしてランダムアクセス可能なメモリーやハードディスクに保存される(映像ソフト蓄積部13)。デジタル録画したオリジナル映像をセグメントリスト、エッセンスユニットリストで管理することにより、双方のリンクをとることが容易にできる。

1) セグメント

・1秒以下のショットはその前のショットにつなげる。

- ・セグメントは文章の文節に相当するもので、画像の切れ目（カット）、セリフの切れ目、テロップの切れ目等から総合的に判断した、映像の切れ目から切れ目迄とする。

- ・オリジナル映像はセグメントリストで管理される。
- ・全てのセグメントをつなぎ合わせれば、オリジナル映像となる。

- ・セグメントをまたぐセリフ、テロップはない。（セリフ、テロップがカットをまたいだ場合、そのカットでセグメントを分けない）

【0043】2) セグメントリスト

- ・セグメントリストは、セグメントの通しNo.、セグメントの先頭フレームNo.、フレーム数とする。

3) エッセンスユニット

- ・セグメント毎にエッセンスユニットを作成する。
- ・エッセンスユニットは、セグメントの中から効果音、テロップ、セリフ部分を中心に切り出した映像である。
- ・1つのセグメントから1つのエッセンスユニットを作成する。

- ・エッセンスユニットの始めと終わり0.5秒は音を絞る。（音声をスムーズにつなぐ）

【0044】4) エッセンスユニット評価

- ・タイトルと終わり(END)だと思われる画像の評価点は高くする。

- ・オープニング、エンディングは他と比べ評価点が高くなるようにする。

- ・エッセンスユニットの長さにより点数を変える。
- ・各エッセンスユニットの中に含まれている要素により、エッセンスユニットに点数をつける。

- ・高得点のエッセンスユニットが連続しないよう、調整する。

5) エッセンスユニットリスト

- ・エッセンスユニットリストは、エッセンスユニットの通しNo.、先頭フレームNo.、フレーム数、評価ポイント等で構成される。

6) セグメント／エッセンスユニット対応リスト

- ・セグメントとエッセンスユニットをリンクさせるために使用される。

(3) 再生

1) プレイ時間設定

- ・VEプレイする場合、最初にプレイ時間（または圧縮率）の設定をする。

2) エッセンスユニットの選択

- ・設定された時間に応じて、点数の高いエッセンスユニットを選択する。

- ・選択したエッセンスユニットの総時間が設定されたプレイ時間と同じにならない場合、実際の再生時間（エッセンスユニットの総時間）を少しだけ長くする。

3) エッセンスユニットの接続・再生

- ・選択したエッセンスユニットをNo.順に接続する。

- ・エッセンスユニット接続しながら、再生を行う。

4) VE→オリジナル、オリジナル→VEの切替

- ・VE再生中に、「VE→オリジナル」の操作が行われた場合、再生中のエッセンスユニットとリンクされているセグメントへ飛び、そこからオリジナル映像を再生する。

- ・オリジナル再生中に、「オリジナル→VE」の操作が行われた場合、再生中のセグメントとリンクされているエッセンスユニットへ飛び、そこからVE映像を再生する。

5) スキップ

- ・オリジナル再生中にスキップボタンを押すと、セグメント毎に映像をとばして見られる。

- ・VE再生中にスキップボタンを押すと、エッセンスユニット毎にVE映像をとばして見られる。

(4) セグメント

1) セグメントの作成方法

- ・カットの位置を検出（判別）する。

- ・1秒以下のショットは前のショットにつなげ1ショットとして処理する。

- ・ショットとショットに効果音、テロップ又はセリフがまたがっている場合は、ショットをつなげて1つのショットとみなす（図5）。

2) セグメントリスト

セグメント通しNo.、先頭フレームNo.、各セグメントを構成するフレーム数を表にする（表1）。

(5) エッセンスユニット

1) エッセンスユニットの形成方法

- ・1つのセグメントから1つのエッセンスユニットを作る。

- ・セグメント中でセリフやテロップのある部分を優先的にエッセンスユニットとして抜き出す。

- ・セリフよりエッセンスユニットを長くしセリフが途切れないようにする。

- ・長いテロップは2秒でカット。

- ・エッセンスユニット化する場合の映像要素は、効果音、テロップ、セリフ又は人物（対象物）とする（図6）。エッセンスユニットリストの例は表2に示す。

(5) インデックスについて

1) インデックス画面の作成方法

- ・各セグメントの最初のフレーム画像を小型化する。

- ・小型化した画像（サムネイル画像）をインデックス画面に張り付ける（図12）。

- ・サムネイル画像とそれに該当するオリジナル映像間にリンクを張る。

- ・サムネイル画像がクリックされたときオリジナル映像を再生する。

【0045】6. 要約生成フロー

本実施の形態例での要約生成手順を図7～図9に示す。

1) 録画・VE生成処理（図7）

・録画操作 ……アナログ映像（TV放送／ビデオ）を入力し録画ボタンをクリック。パソコン+ビデオボードによりAVIファイルを生成。

・オリジナル映像…デジタル録画し各フレームにNo.を付けランダムアクセス可能なファイルとして保存（AVIファイル）。

・カット判別 ……カットからカットまで（=ショット）を抽出。

【0046】・セグメント編集 ……1秒以下のショットやカットをまたぐセリフ／テロップのあるショットをつなげ、セグメントにする。

・リスト生成 ……セグメントリストを生成する（表1）。このリストは、セグメント開始フレームNo.とフレーム数を記録したものである。

・エッセンスユニット(EU) ……映像ソフト構成要素（効果音、テロップ、セリフ、対象物）の存在状態に基いてエッセンスユニットを作る。エッセンスユニットはセグメントの1部分である（1セグメントから1EUを生成）。

・EU評価 ……EU中の要素（効果音、テロップ、セリフ、対象物）により重要度を点数付けを行なう。EUリスト生成とサムネイル画面抽出を行なう。EUリストには、EU開始・終了フレームNo.、フレーム数、重要度などを記録する（表2）。また、各EUの最初の静止画1枚をサムネル化し、インデックス画面（図12）を生成する。これら処理により、オリジナル映像=Σセグメント映像>ΣEU映像となる。

【0047】2) VEプレイ（図8）

要約再生は以下の如く実行される。

・タイトル選択 ……プレイ画面（図11）で「HOME」キーをクリック。ライブラリーパネル（図10）が表示される。この画面では、要約処理済みの映像ソフトのタイトル画とタイトルが表示されるので、所望のタイトルまたはタイトル画をクリック。

・VEプレイ ……プレイ画面でVEをクリック。

・プレイ時間／要約率設定 ……プレイ画面でVE設定バー31をドラッグ。

・EU選択 ……設定されたプレイ時間に合うように、全EUリストから重要度の高いEUから選択。

【0048】・EUリスト ……VE再生のため、EUリストを参照する。

・VE再生 ……VE・EUリストのフレームNo.より、オリジナル映像を再生する。

・VE→オリジナル／オリジナル→VE ……VE・EUリストとリンクしているセグメントリストの映像、又はセグメントリストとリンクしているVE・EUリストの映像を再生する。

・スキップ ……オリジナルプレイ時には再生中の次のセグメントから、VE再生時には再生中の次のEUからそれぞれ再生する。

【0049】3) インデックスからの映像呼び出し（図

9)

・INDEX画面呼び出し ……プレイ画面（図11）でINDEXボタンをクリックする。図12のINDEX画面が表示される。

・見たいサムネイル画像を選択 ……視聴者が所望するサムネイル画像をクリックする。

・サムネイル画像に対応するオリジナル映像を再生…クリックされたサムネイル画像と対応するセグメントが再生される。再生はオリジナルモード（通常の再生形式）の映像となる。

・VEプレイ ……図11のプレイ画面でVEをクリックする。

・プレイ時間／要約率設定 ……プレイ画面でVE設定バー31をドラッグする。これで要約の再生時間が設定される。

【0050】・EU選択 ……各エッセンスユニットの再生時間の合計が、この設定された要約再生時間の合計に達するまで、エッセンスユニット評価リスト（表2）の中から点数の高いEUを順に選択。

・EUリスト ……当該VE再生のためのリストを作成する。同じ映像ソフトについてのものであっても、要約再生時間は視聴者の設定如何でその都度異なる。このEUリストは要約再生の都度作成される。

・VE再生 ……EUリストのフレームNo.より、オリジナル映像を再生。

【0051】7. 操作画面

VEの録画及び再生の操作は各画面表示とマウスとで行なわれる。これらは、主として録画及び録画されたライブラリーを表示するライブラリーパネル（図10）と、VEやオリジナル映像の再生のための操作を行うプレイパネル（図11）と、サムネイル画面を表示するINDEX画面（図12）とで構成される。

【0052】(1) ライブラリーパネル（図10）

映像ソフト蓄積部13に収録されている全映像のファイル管理を行う。ユーザーはこのパネルで、見たい映像の検索や呼び出し操作、デジタル映像データからVE生成操作、VEプレイ画面の呼び出し操作、インデックス画面呼び出し操作等を行う。

・ライブラリー表示（図10）

ここでは、映像ソフト蓄積部に保存されている映像の代表的静止画像（タイトル画像）を表示を行なう。ファイル名（タイトル）の入力／表示。ライブラリーを選択してプレイ、インデックス表示、消去。

・タイトル画像

もっとも重要度の高いエッセンスユニットの最初の画像1枚をサムネイル化しそのタイトルの代表的画像として表示する。

・タイトル入力ウィンドウ

タイトル画像の横／下部にタイトルを手入力するためのウィンドウを表示する。

・VE生成ボタン

オリジナル映像からVE生成、インデックスの生成を行う。VE生成、インデックス生成が処理されたタイトル画像には、VEマークが表示される。

・各タイトル画像／タイトル名をクリックすることにより、プレイパネルへ画面が変わりクリックされたタイトルが選択表示される。

【0053】(2) プレイ画面(図11)

ライブラリーパネルで選択されたタイトルの、オリジナル映像やVE映像を再生したり、インデックス画面を呼び出すためのパネルである。

・ライブラリーボタン

ライブラリーパネル(図10)へ戻る。なお図11では「HOME」と表示されている。以下の記述での鉤括弧書きもそのボタンの画面表示を表わす。

・インデックスボタン「INDEX」

プレイ中の映像のインデックス画面を表示する。

・VE／オリジナル切替ボタン「VE/OR」

VEプレイ中にオリジナルボタンを押すと、対応するオリジナル映像を再生する。オリジナル再生中にVEボタンを押すと、対応するVE映像を再生する。VEプレイ中にVE時間を再設定する場合に押す。

【0054】ここで、VEを再生中にオリジナル映像に切り替える場合、再生中のエッセンスユニットに対応するセグメントNO.を対応リストから、さらにそのセグメント開始フレームNO.をセグメントリストから探し、オリジナル映像の再生するフレームNO.を決定する。オリジナル映像を再生中にVEに切り替える場合、再生中のセグメントに対応するエッセンスユニットの開始フレームからエッセンスユニットリストに従って再生する。

【0055】・スタート／停止ボタン「START/STOP」

VE又はオリジナル再生のスタート／停止

・スキップ送り／戻りボタン「<SKIP>」

再生中の映像のスキップ送り及びスキップ戻り。

【0056】ここで、VEを再生中にスキップ送りボタンが操作された場合、再生中のエッセンスユニットの次のエッセンスユニットの開始フレームから再生する。スキップ戻りボタンが操作された場合は再生中のエッセンスユニットの1つ前のエッセンスユニットの開始フレームから再生する。オリジナル映像を再生中にスキップ送りボタンが操作された場合、再生中のセグメントの次のセグメントから再生する。スキップ戻りボタンが操作された場合は再生中のセグメントの1つ前のセグメントの開始フレームから再生する。

【0057】・VE時間設定／プレイ位置表示バー(符号31)

「設定」ボタンを押してこのバーをドラッグし、プレイ時間の設定を行う。この操作により、ユーザの好きな時間で映像の要約を見ることができる。VE又はオリジナ

ルの再生中には再生位置の表示を行う。(これにより、ユーザは残りの映像を見るための時間が一目でわかる)

・タイトル情報「TITLE」

再生中のタイトル(ファイル名)を表示する。

【0058】(3) インデックス画面(図12)

プレイパネルでプレイ中のタイトルのインデックス画像を表示する。インデックス画面に並んでいる各サムネイル画面で、見たいシーンを検索し、呼び出すことができる。

・タイトル名

手で入力されたタイトルのテキストが表示される。

・サムネイル画像

各エッセンスユニットの最初の画像をサムネイル化した画像。これをクリックするとプレイパネルへ戻り、クリックされた映像からプレイすることができる。

【0059】以下、変形例について説明する。本発明はテレビジョン受像機、ビデオテープレコーダー、DVD／レーザディスクプレーヤその他、種々のオーディオビデオ関連の家庭電化製品に組み込むことが出来る。アプリケーションプログラムの形にしてパーソナルコンピュータ上でも実施出来る。この場合はCD-ROMに格納して配布したり、ネットワークを介して配布することになる。映像ソフトについて、製作側に於てセグメント、エッセンスユニット、その評価点等を求め、これを映像ソフトに付加してDVD等に格納し、視聴者側で、この付加情報に基いて設定された再生時間に合致するようにエッセンスユニットを選択して再生するようにも良い。映像ソフトの入手経路は例示した4つに限られない。例えばケーブルテレビジョンその他、例示しない種々の形態での入手が考えられる。

【0060】

【発明の効果】以上説明したように、本発明では映像の各カット位置を検出し、各カット位置から次のカット位置までの各ショットについて、その持続時間が所定値以下のもの、或いはセリフ、文字その他注目した映像ソフト構成要素が両者間に跨がっているものを連結し、単数又は複数のショットから成る各セグメントを形成し、効果音、文字、セリフ、対象物その他注目した映像ソフト構成要素の存在状態に対応して、各セグメントの中から所要長さの映像をエッセンスユニットとして取り出すようにした(請求項1, 3)。従って、先の出願に於けるショット単位での代表画像抽出に比べ、一層的確な形での要約表現が可能になった。

【0061】更に、このようにして抽出した各エッセンスユニットについて、その抽出機とされた映像ソフト構成要素の存在状態に対応した評価点を付与し、この評価点を基に、指定された再生時間になるように前記エッセンスユニットの中の幾つかを選択し、これらを再生するようにした(請求項2, 4)。従って、一層内容が的確になった上記の如き要約を、夫々の視聴者がそれに振

り向けられる時間に応じて見ることが出来、限られた時間を有効活用して、洪水の如く供給される映像ソフトの中から、各視聴者が真に所望するものを、迅速且つ的確に選び出すことが出来る。

【図面の簡単な説明】

【図1】映像ソフトの構成の一例（詳細）を示す図。

【図2】映像ソフトの構成の一例（概略）を示す図。

【図3】実施の形態のVEプレーヤの構成を示すブロック図。

【図4】音声出力の一例を示すグラフ。

【図5】ショットとセグメントとの関係例を示す図。

【図6】セグメントとエッセンスユニットとの関係例を示す図。

【図7】録画・VE生成時の処理手順例を示すフローチャート。

【図8】VE再生時の処理手順例を示すフローチャート。

【図9】インデックスからの映像呼び出しの処理手順例を示すフローチャート。

【図10】ライブラリーパネル（HOME画面）の例を示す正面図。

【図11】プレイ画面の例を示す正面図。

【図12】INDEX画面の例を示す正面図。

【図13】映像データの差分とカット位置の関係例を示すグラフ。

【図14】Y（輝度）分布の差分の例を示すグラフ。

【符号の説明】

1 1…TV受信部	1 2…入力制御部
1 3…映像ソフト蓄積部	1 4ビデオデータ
1 6…光ディスクプレーヤネット接続部	1 7…インターフェース
1 8…カット判別部	1 9…データ蓄積部
2 1…要素検出部	2 2…セグメント
2 3…エッセンスユニット抽出部	2 4…エッセンスユニット評価部
2 6…ユーザインタフェース	2 7…ディスプレイ
2 8…再生制御部	2 9…スピーカ
3 1…VE設定パレーヤ。	1 0 0…VEブ

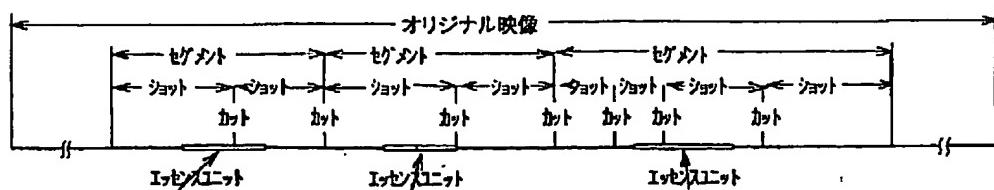
【表1】

セグメントNo.	1	2	3	4	5	6	7	8……
先頭フレームNo.	1	61	121	201	301	391	511	621……
フレーム数	60	60	80	100	90	120	110	……

【表2】

エッセンスユニットNo.	1	2	3	4	5	6	7	8……
先頭フレームNo.	20	65	130	220	320	400	520	621……
フレーム数	40	50	60	60	60	90	90	……
評価ポイント	10	13	15	10	9	8	10	……

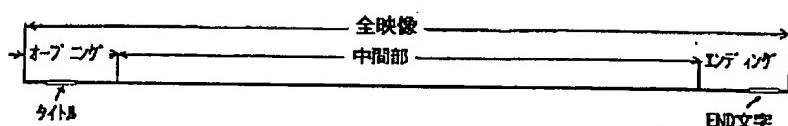
【図1】



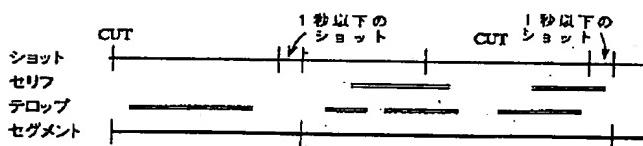
(11)

特開平11-220689

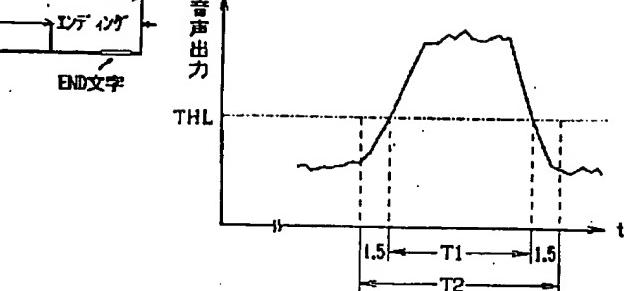
【図2】



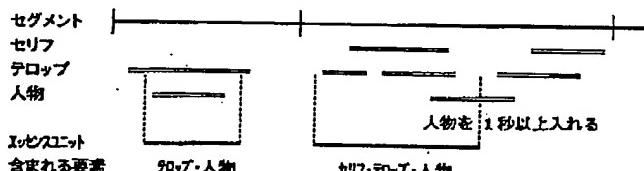
[☒ 5]



【図10】

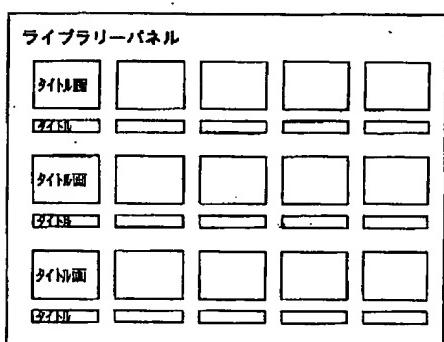


[图 6]

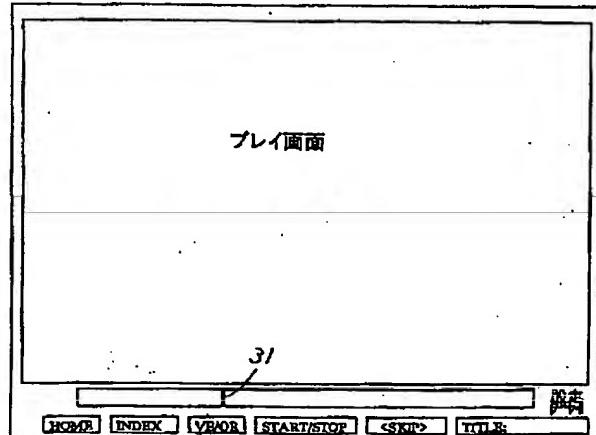


【図11】

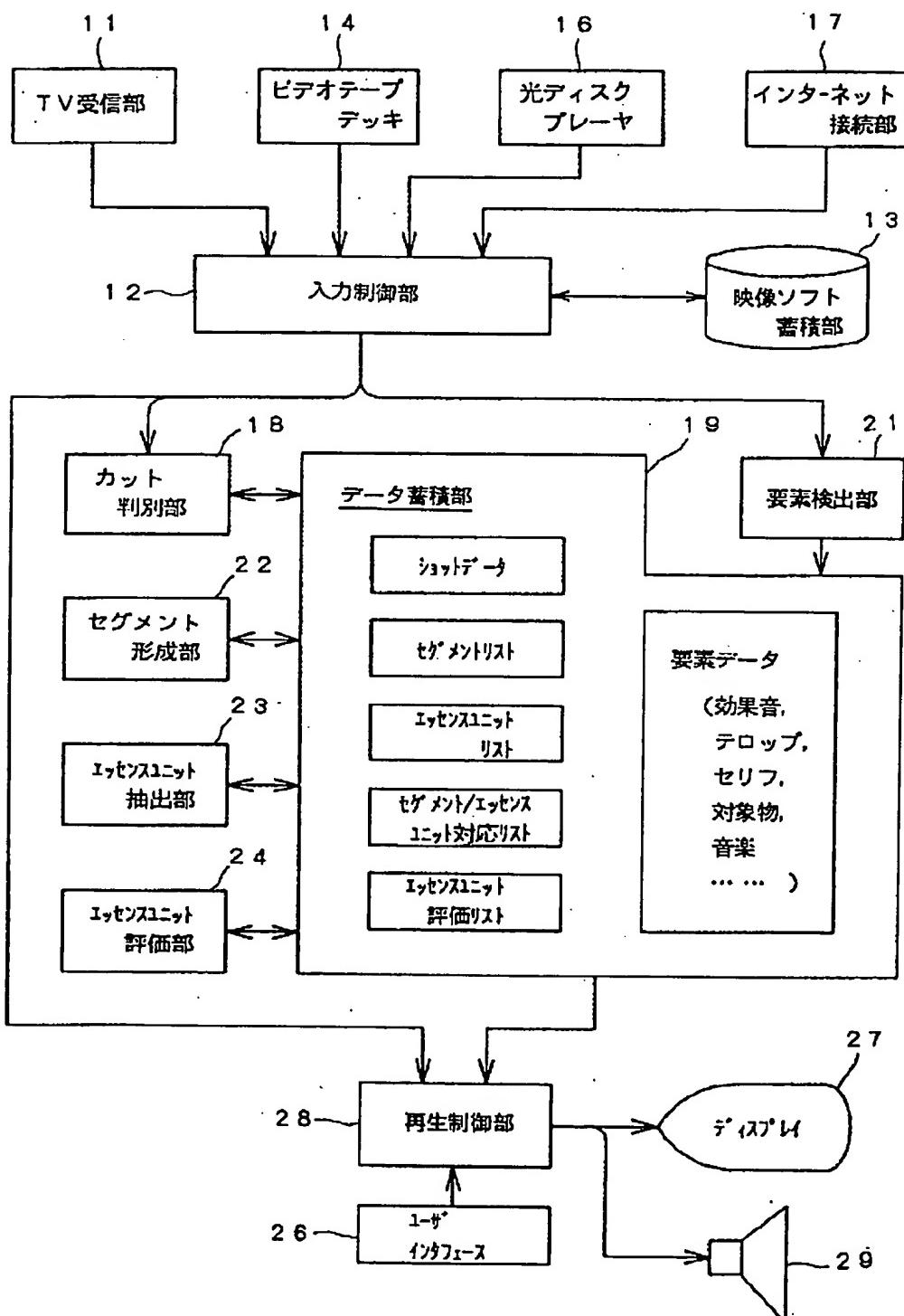
(图 1-2)



プレイ画面

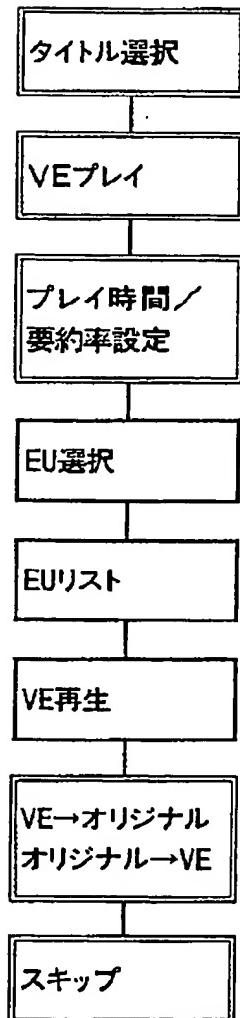


【図3】

100 VEプレーヤ

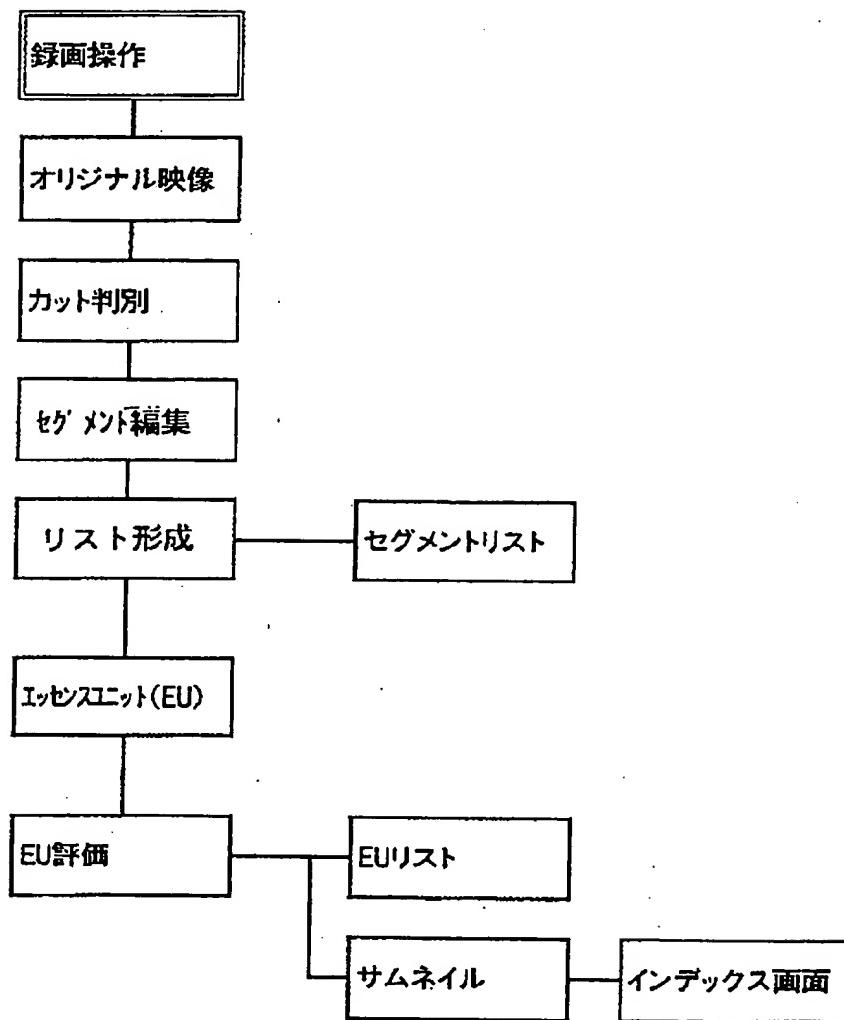
【図8】

VEプレイ

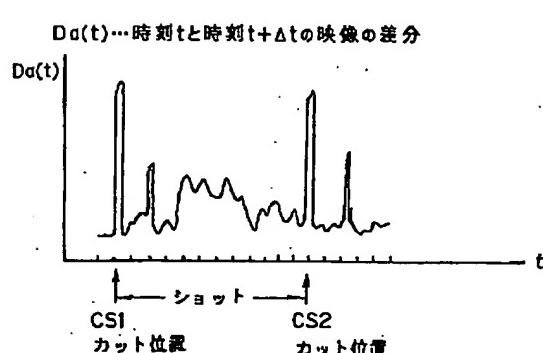


【図7】

録画・VE生成

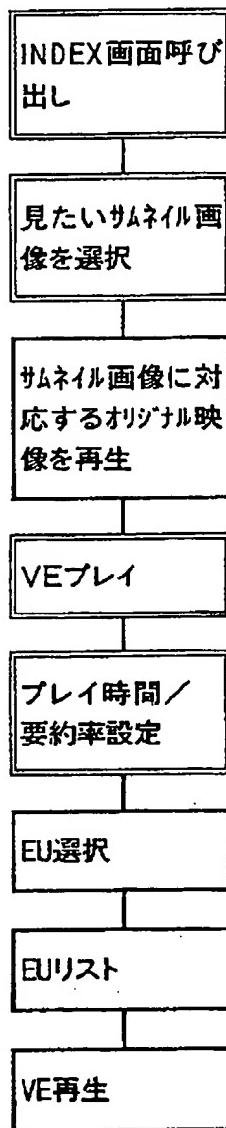


【図13】



【図9】

インデックスから映像呼び出し



【図14】

